Maximum Likelihood Estimation of the Negative Binomial Distribution

11-19-2012

Stephen Crowley
stephen.crowley@hushmail.com

Abstract. Maximum likelihood estimation of the negative binomial distribution via numerical methods is discussed.

1. Probabilty Function

1.1. Definition.

The probability density function(pdf) of the (discrete) negative binomial(NB) distribution[3] is given by

$$p_{\mathrm{nb}}(y|r,p) = \begin{cases} 0 & y < 0 \\ \frac{\Gamma(r+y)p^r(1-p)^y}{\Gamma(r)\Gamma(y+1)} & y \geqslant 0 \end{cases} \tag{1}$$

where the notation $y|r,p$ means "y given r and p" with r and p being parameters of the density function and y being the outcome variable. r is the number of events until the experiment is stopped and p is probabilty of success in each event. Since the NB is a discrete distribution we have

$$\begin{aligned} \sum_{y=0}^{\infty} p_{\mathrm{nb}}(y|r,p) &= \sum_{y=0}^{\infty} \frac{\Gamma(r+y)p^r(1-p)^y}{\Gamma(r)\Gamma(y+1)} \\ &= 1 \end{aligned} \tag{2}$$

1.1.1. Moments.

The mean of the NB distribution is

$$\sum_{y=0}^{\infty} y p_{\mathrm{nb}}(y|r,p) = \frac{r(1-p)}{p} \tag{3}$$

and the variance

$$\sum_{y=0}^{\infty} \left(y - \frac{r(1-p)}{p}\right)^2 p_{\mathrm{nb}}(y|r,p) = \frac{r(1-p)}{p^2} \tag{4}$$

Of course the higher moments exist but those will not be listed here.

1.1.2. Log-Likelihood Function.

The log-likelihood function is the same function as the logarithm of the probability density, just considered from a different perspective.

$$\begin{aligned} \ell(r,p|y) &= \ln(p_{\mathrm{nb}}(r,p|y)) \\ &= \ln(p_{\mathrm{nb}}(y|r,p)) \\ &= \ln(\Gamma(r+y)p^r(1-p)^y) - \ln(\Gamma(r)\Gamma(y+1)) \end{aligned} \tag{5}$$

It is more convient to work with the log-likelihood function $\ln(p_{\mathrm{nb}}(r,\, p|y))$ than the likelihood function $p_{\mathrm{nb}}(r,p|y)$ and to consider the value of the parameters r, p given the vector of observed data $y_1, y_2, \ldots, y_N$.

$$\begin{aligned} \ell(r,p|y_{1\ldots N}) &= \sum_{i=1}^{N} \ell(r,p|y_i) \\ &= Nr\ln(p) - \mathrm{N}\ln(\Gamma(r)) + \sum_{i=1}^{N} \ln(\Gamma(r+y_i)) + y_i \ln(1-p) - \ln(\Gamma(y_1+1)) \end{aligned} \tag{6}$$

2. Maximum Likelihood Estimation

2.1. Derivatives.

To find the maximum likelihood estimates of r and p given $y_{1\ldots N}$ we first differentiate $\ell(r, p|y_{1\ldots N})$ with respect to p and set it to 0.

$$\frac{\partial\ell(r,p|y_i)}{\partial p} = \frac{Nr}{p} + \sum_{i=1}^{N} -\frac{y_i}{1-p} = 0 \tag{7}$$

to get the maximum likelihood estimate of p

$$p = \frac{Nr}{Nr + \sum_{i=1}^{N} y_i} \tag{8}$$

Then differentiate with respect to r and substitute the maximum likelihood estimate of p to eliminate it from the equation and get an equation with 1 unknown instead of 2.

$$\begin{aligned}\frac{\partial\ell(r,p|y_i)}{\partial r} &= N\ln(p) - N\Psi(r) + \sum_{i=1}^{N} \Psi(r+y_i) \\ \frac{\mathrm{d}\ell(r|y_i)}{\mathrm{d}r} &= N\ln\left(\frac{Nr}{Nr+\sum_{i=1}^{N} y_i}\right) - N\Psi(r) + \sum_{i=1}^{N} \Psi(r+y_i) \\ &= 0\end{aligned} \tag{9}$$

The equation $\frac{\mathrm{d}\ell(r|y_i)}{\mathrm{d}r}$ has no closed-form solution so the root has to be found with numerical methods.

2.2. Estimation.

2.2.1. Brent's Algorithm.

Brent's algorithm[1] is particularly well-suited to the optimization of this problem and an implementation of the algorithm in the Java language is available as part of the Apache Commons Mathematics Library[2].

Bibliography

[1] R.P. Brent. *Algorithms for minimization without derivatives*. Dover Publications, 2002.

[2] The Apache Software Foundation. The apache commons mathematics library. Https://commons.apache.org/math/.

[3] J.M. Hilbe. *Negative binomial regression*. Cambridge University Press, 2011.